

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) EP 0 955 759 A1

(12)

## **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:10.11.1999 Bulletin 1999/45

(51) Int Cl.6: H04L 29/08

(21) Numéro de dépôt: 99401074.2

(22) Date de dépôt: 03.05.1999

(84) Etats contractants désignés:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU

MC NL PT SE

Etats d'extension désignés:

AL LT LV MK RO SI

(30) Priorité: 06.05.1998 FR 9805717

(71) Demandeur: FRANCE TELECOM 75015 Paris (FR)

(72) Inventeurs:

 Bonjour, Dominique 22300 Lannion (FR)

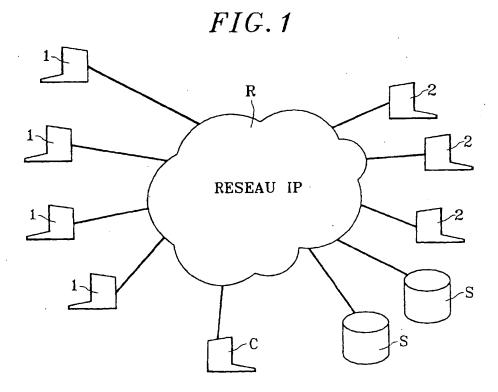
Rabret, Laurent
 22300 Lannion (FR)

(74) Mandataire: Loisel, Bertrand
 Cabinet Plasseraud,
 84, rue d'Amsterdam
 75440 Paris Cédex 09 (FR)

# (54) Procédé et dispositif pour évaluer les consultations de sites web

(57) On équipe de moyens de mesure un ensemble de terminaux informatiques (1) reliés au réseau (R). Les moyens de mesure (7) dont sont équipés un terminal de cet ensemble reçoivent des messages d'accès émis par le terminal vers le réseau, retransmettent ces messages sur le réseau, détectent si chacun de ces messages

concerne l'un des sites dont les consultations sont à évaluer, et adressent des données de mesure à un serveur de collecte relié au réseau en réponse à la détection d'au moins un message concernant l'un des sites en question. Le serveur de collecte assure un traitement statistique des données de mesures reçues pour estimer les nombres de consultations des sites.



30

#### Description

[0001] La présente invention concerne le domaine des réseaux informatiques, notamment de l'Internet, basés sur la famille de protocoles IP (voir D. Comer, "Internetworking with TCP/IP", Prentice Hall, 1995). Elle vise à offrir une solution au problème de la mesure fiable et neutre de l'audience des différents services et programmes disponibles sur un tel réseau. En particulier, la caractérisation du flux de requêtes envoyées vers un serveur Web particulier est souhaitable, de même que celle de l'écoute de canaux de distribution (multicast).

1

[0002] L'invention a pour application la mise en place d'une infrastructure de suivi de l'audience des services proposés sur des réseaux tels que l'Internet, permettant des traitements statistiques selon des principes analogues à ceux utilisés aujourd'hui pour la mesure de l'audience des chaînes de télévision. L'existence de mesures d'audience fiables et neutres est un facteur nécessaire pour faire de l'Internet un media grand public, crédible auprès des annonceurs.

[0003] Une famille d'outils et de méthodologies disponibles aujourd'hui vise à mesurer l'audience au niveau des serveurs (voir par exemple US-A-5 732 218). Ceci permet d'obtenir un comptage des requêtes de pages Web hébergées par ce serveur. Les suivis d'audience sont actuellement réalisés sur la base de ces indications

[0004] Une amélioration de la méthode précédente, en cours de développement, vise à rendre plus représentatives les mesures fournies en prenant en compte l'action des "caches" (qui masquent certaines requêtes). Cette étude est en cours dans le cadre de l'IETF (voir J. Mogul, P. Leach; "Simple hit metering and usage limiting for HTTP", Draft RFC, Internet Engineering Task Force, 1997; ou J. Pitkow, « In search of reliable usage data on the WWW », Computer Network and ISDN systems, Vol. 29, No. 8-13, septembre 1997, pages 1343-1355). Elle consiste pour les "caches" à aller informer les serveurs du nombre de réponses qu'ils ont 40 fournies en leur nom.

[0005] Aucune solution connue ne propose de mesurer l'audience des services réseau au niveau des terminaux d'utilisateurs. La mesure d'audience au niveau des serveurs souffre de limitations multiples:

- elle nécessite d'équiper chaque serveur à cet effet, ce qui est difficile à gérer compte tenu de leur prolifération;
- elle n'est pas adaptée à la fourniture de mesures d'audience par un tiers indépendant;
- elle est peu fiable à cause des mécanismes de "cache" et d'agents dits "proxy" (qui envoient des requêtes au nom d'un ensemble de terminaux, souvent pour des raisons de sécurité);
- elle n'est pas applicable aux techniques de distribution (multicast).

[0006] Les études sur la remontée par les "caches" du nombre de requêtes qu'ils ont interceptées ne s'intéressent qu'à la troisième limitation, et élargissent le problème à un nouvel ensemble d'équipements: à moins de contrôler tous les "caches", on ne peut pas être certain de la véracité de leurs informations, et il est donc difficile de considérer ces chiffres comme fiables. [0007] Dans W098/10349, il est proposé que le site consulté retourne la page Web demandée avec un morceau de programme qui, chez le client, surveille l'affichage à l'écran de la page et commande le renvoi d'informations relatives à cette surveillance. Ce document aborde le problème de la non neutralité du site consulté vis-à-vis de la mesure d'audience: Une solution proposée pour traiter ce problème consiste à crypter les informations de surveillance retournées par le client de façon que seul un tiers indépendant soit capable de les décrypter. Une autre solution consiste en ce que le site consulté réponde à une requête issue d'un client en retransmettant la requête et la page concernée au site indépendant, lequel se charge de retransmettre la page demandée et le morceau de programme approprié vers le site du client. Celui-ci retourne alors les informations de surveillance directement au site indépendant. Les solutions préconisées dans ce document permettent de ne pas équiper les terminaux clients d'un logiciel spécial pour la mesure d'audience. Le but poursuivi est d'essayer que les mesures effectuées soient le plus exhaustives possible. Mais ces solutions impliquent :

- que le contenu des pages résidentes dans le serveur de contenu soit modifié afin d'émettre une requête vers le secteur de collecte en même temps que le contenu intéressant le client (méthode non universelle nécessitant de rééditer l'ensemble des pages sur les serveurs);
- que l'information soit acheminée par un serveur indépendant avant d'arriver chez le client, ce qui a nécessairement un impact sur la qualité perçue du

[0008] L'invention a pour but de permettre une nouvelle approche de la mesure de l'audience sur Internet, basée sur une observation à la source des requêtes produites par les usagers.

[0009] L'invention propose ainsi un procédé tel que défini dans la revendication 1.

[0010] Un panel de clients, dont les terminaux sont équipés des moyens de mesure, pourra alors fournir en temps réel à une entité indépendante gérant le serveur de collecte le compte exact des sites explorés et des canaux écoutés. Le procédé présente les avantages:

de permettre d'utiliser les modèles de mesure d'audience (panel choisi sur la base de sa représentativité sociologique) qui ont été développés pour les programmes de télévision. En particulier, il est possible d'associer un profil d'utilisateur à cha-

55

que terminal observé, ce qui rend plus pertinentes les données obtenues;

- de ne pas nécessiter d'instrumentation spécifique au niveau de chaque serveur ou canal observé, étant donné qu'on se place à la source des requêtes;
- d'être insensible aux interventions des "caches" puisque les requêtes sont observées dès leur source;
- d'être déployable par un acteur économique indépendant des réseaux et des services empruntés, ce qui est capital vis-à-vis de la neutralité des informations fournies;
- de s'appliquer également aux services Internet utilisant la distribution;
- de pouvoir observer en temps réel les actions (requêtes) des usagers du réseau;
- d'être transparent (c'est-à-dire non contraignant et évolutif) vis-à-vis des applications utilisées pour "naviguer" sur le réseau.

[0011] Un mode de réalisation préféré du procédé selon l'invention est énoncé dans la revendication 2.

[0012] De cette manière, on est capable non seulement d'évaluer les nombres de consultations des sites, mais encore d'analyser la qualité offerte aux usagers lors de ces consultations (temps de réponse, vitesse, volumes d'information transmis...). Les résultats de ces analyses peuvent être communiqués aux personnes gérant les sites concernés pour leur permettre d'adapter les moyens qu'elles mettent en oeuvre.

[0013] Un autre aspect de l'invention se rapporte à un module de détection de messages et de transmission d'informations tel que défini dans la revendication 3.

[0014] D'autres particularités et avantages de la présente invention apparaîtront dans la description ciaprès d'exemples de réalisation non limitatifs, en référence aux dessins annexés, dans lesquels:

- les figures 1 et 2 sont des représentations schématiques de réseaux informatiques; et
- la figure 3 est un schéma d'un terminal équipé d'un module selon l'invention, illustrant son fonctionnement avec le réseau.

[0015] La figure 1 montre schématiquement un réseau R fonctionnant selon un protocole de la famille IP ("Internet Protocol"), auquel sont reliés un certain nombre de terminaux 1 depuis lesquels des requêtes d'accès peuvent être émises sur le réseau R. La référence 2 désigne des serveurs hébergeant des sites (pages Web) qui peuvent être consultés par les terminaux 1 via le réseau.

[0016] Un ou plusieurs serveurs de collecte S sont reliés au réseau R afin de recevoir et traiter des données de mesure issues de certains terminaux 1. L'entité gérant un tel serveur de collecte S s'adresse à un panel d'usagers détenteurs de terminaux 1 reliés au réseau

R, et leur fournit des moyens de mesure consistant en un module 7 décrit ci-après. Les modules 7 interceptent les messages de requête d'accès émis par des applications exécutées dans les terminaux en question, pour identifier les sites interrogés, déterminer le type de navigateur utilisé et effectuer des mesures dont les résultats sont ensuite fournis au serveur S. Celui-ci peut ensuite effectuer toutes sortes de traitements statistiques pour analyser l'audience des sites 2. Le panel d'usagers est donc déterminé en fonction de critères de représentativité du même genre que ceux utilisés dans le domaine de la mesure d'audience des chaînes de télévision. [0017] Dans le cas où il y a plusieurs serveurs de collecte S, un serveur de configuration C est accessible par le réseau R afin de fournir aux modules 7 les informations de localisation (telles que l'adresse, le numéro de port et le protocole à utiliser) du serveur de collecte S auquel ils doivent s'adresser.

[0018] Les applications informatiques conçues pour accéder au réseau utilisent des interfaces logicielles de programmation ("Application Programming Interface", ou API). Ces interfaces de communication en réseau font partie des ressources offertes par le système d'exploitation de chaque terminal. Elles permettent d'accéder aux protocoles de communication implantés en utilisant un ensemble de "primitives", qui appellent des fonctions élémentaires permettant d'ouvrir une connexion vers une adresse distante, d'envoyer et de recevoir des données, de sélectionner un canal de distribution (multicast channel), etc. Ces primitives sont ensuite traduites en messages élémentaires du protocole en question (voir illustration en figure 2). Dans les terminaux, le fonctionnement selon le protocole de communication est assuré par d'autres ressources logicielles du système d'exploitation (gestionnaire de protocole) associées à des cartes de communication appròpriées. [0019] La figure 3 montre un module 7 de détection de messages et de transmission d'informations dont sont équipés les terminaux 1, raccordés au réseau R, des usagers du panel retenu par l'entité gérant un serveur de collecte S.

w E

. 3

- 3

. . .

[0020] Sur la figure 3, la référence 5 désigne les éléments de la couche de protocole IP dont est équipé un terminal 1, et la référence A désigne l'API correspondante, prévue dans le terminal 1 pour réaliser l'interface entre le gestionnaire de protocole 5 et les applications 6 exécutées sur le terminal 1 et susceptibles de communiquer via le réseau.

[0021] Le module logiciel 7 présente vis-à-vis des applications 6 une interface notée A' qui comprend au minimum l'ensemble obligatoire des fonctions offertes par l'interface A:

[0022] Dans le cas courant où le terminal fonctionne avec le système d'exploitation Windows commercialisé par la société Microsoft, l'interface A peut être l'interface Winsock. Dans ce cas, le module logiciel 7, qui présente l'interface A vis-à-vis des applications 6, porte le nom habituel du gestionnaire de protocole standard

(Wsock32.dll dans le cas des versions Windows 95 et Windows NT 3.5 ou 4.0). Le module 7 appelle à son tour le gestionnaire de protocole standard 5 qui aura été isolé des applications 6, c'est-à-dire placé dans un répertoire non prioritaire.

[0023] Le module 7 comprend un filtre de primitives 8, un décodeur de requêtes 9, un module 10 de gestion de la configuration, et un module 11 de transmission de données.

[0024] Le filtre de primitives 8 reçoit les messages conformes à l'interface A. Si le message est issu d'une application 6 et correspond à un flux de données sur lequel l'audience doit être mesurée, il est fourni au décodeur de requêtes 9. Sinon, il est relayé de façon inchangée vers le gestionnaire de protocole 5. Ceci permet d'assurer la transparence du procédé, les messages n'étant pas altérés. La définition et l'identification des flux à observer sont fournis au filtre de primitives 8 par le module de configuration 10.

[0025] Le décodeur de requêtes 9 agit sur les flux aiguillés par le filtre de primitives 8. Il analyse le contenu des messages sélectionnés et en déduit des enregistrements élémentaires, ou données de mesure, constitués des informations d'audience (serveur consulté, canal écouté) auxquelles est attachée une étiquette d'horodatage. Les messages analysés sont eux aussi relayés, de façon inchangée, vers le gestionnaire de protocole 5.

[0026] Le module de gestion de la configuration 10 permet de programmer le filtre 8 et le décodeur 9 en fonction des informations souhaitées. Ce module 10 dialogue avec le serveur de collecte S et/ou avec le serveur de configuration C via le gestionnaire de protocole 5 et le réseau R, ce qui permet de faire évoluer aisément sa configuration.

[0027] Le module de transmission des données 11 dialogue avec le serveur S via le gestionnaire de protocole 5 et le réseau R, en lui renvoyant en temps réel les enregistrements élémentaires produits par le décodeur

[0028] Dès que l'application 6 accède pour la première fois au réseau R, le module de gestion de configuration 10 envoie une requête au serveur de configuration C à travers l'interface A, pour initialiser un dialogue représenté par la référence 20 sur la figure 3. Cette requête inclut au minimum l'identification de l'application 6. Le serveur C retourne des informations 21 caractérisant les flux devant faire l'objet d'observations par le module 7 (ces informations 21 sont transmises au filtre de primitives 8) ainsi qu'une description des informations à collecter (ces informations 22 sont transmises au décodeur de requêtes 9). Les paramètres fournis peuvent notamment inclure:

- la description des applications 6 dont les requêtes devront être décodées;
- l'adresse IP et le port TCP (Transmission Control Protocol) du serveur de collecte S auquel les don-

- nées de mesure devront être transmises:
- l'adresse IP et le port TCP du serveur de configuration C qui servira à la prochaine mise en service du module 7;
- une information permettant de déterminer le décalage entre l'horloge locale (non modifiable) du terminal 1 et l'horloge du serveur S.

[0029] A chaque accès au réseau R depuis le terminal 1, l'application 6 fournit au module 7 un message 25 conforme à la spécification de l'interface A. Le filtre de primitives 8 détermine si ce message 25 doit être pris en considération ou non dans le cadre des mesures d'audience. Si le message n'est pas pris en considération, il est simplement transmis au gestionnaire de protocole 5 (référence 26 sur la figure 3) qui va procéder à son acheminement sur le réseau R jusqu'à sa destination 2 (référence 28). Si le message doit être analysé (référence 27), le décodeur de requêtes 9 extrait de son contenu les informations pertinentes, et en déduit des données de mesure qu'il enregistre. Pour assurer la transparence du module 7, ce message 27 est retransmis par le décodeur de requêtes 9 vers le gestionnaire de protocole 5 pour son acheminement inchangé dans le réseau

[0030] Dans le sens de la réception, tous les messages 32 passés au module 7 par le gestionnaire de protocole 5 sont délivrés sans changement à l'application destinataire 6 via le décodeur de requêtes 9 et le filtre de primitives 8. Le décodeur de requêtes 9 se charge d'estimer des éléments de qualité de service de la connexion. Des exemples de paramètres de qualité de service pouvant être mesurés sont:

- l'accessibilité du serveur distant interrogé (succès de la connexion);
  - le temps de connexion au serveur;
  - la durée nécessaire pour charger le document demandé (fonction du débit utile moyen de la connexion);
  - l'éventuel arrêt du transfert en cours, soit du côté du terminal soit du côté du réseau ...

[0031] Les données de mesure 30 obtenues par le décodeur de requêtes 9 sont fournies au module de transmission de données 11. Ce module 11 se connecte au serveur de collecte S à travers l'interface A et le gestionnaire de prolocole 5, et lui transmet en temps réel les données de mesure produites par le décodeur 9 (message 31).

[0032] Dans le cas du service Web porté par la pile de protocoles TCP/HTTP (voir D. Comer, Internetworking with TCP/IP", Prentice Hall, 1995), le décodeur de requêtes 9 extrait des messages HTTP (Hypertext Transfer Protocol, voir HTTP/1.0, RFC 1945, Internet Engineering Task Force, 1996) les informations d'identification du serveur S, le type de transaction effectué et la nature des documents demandés.

40

30

40

[0033] Par exemple, avec la version 1.0 de HTTP, trois transactions élémentaires sont prévues, à savoir GET (pour demander un document), HEAD (pour demander uniquement l'entête d'un document) et POST (pour transmettre des informations au serveur). Pour chacune de ces commandes, le message est analysé par le décodeur de requêtes 9 afin d'en extraire l'URL (Uniform Ressource Locator) qui identifie à la fois un document et une machine serveur.

[0034] Un traitement spécifique est effectué dans le cas du protocole IGMP (Internet Group Membership Protocol) qui permet aux terminaux de s'inscrire comme membres d'un groupe de distribution particulier (commandé par des options particulières de l'interface A).
[0035] Les données de mesure 30 fournies par le décodeur de requêtes 9 au module de transmission de données 11 peuvent comprendre, sous la forme d'une chaîne de caractères à un format déterminé, les quatre informations suivantes:

- identification du terminal 1, créée lors de l'installation du module 7 dans le terminal 1. Cette identification, en principe invariable, est par exemple codée sur 32 bits. Elle permettra au serveur S d'établir des statistiques en fonction du type d'usagers;
- nom de l'application 6 d'où provient la requête, cette information permettant notamment de connaître le logiciel de recherche employé;
- l'instant d'émission et de conclusion de la requête, permettant d'analyser la distribution temporelle du trafic ainsi que les éléments de qualité de service;
- l'URL ou l'adresse de groupe permettant d'identifier les données demandées.

[0036] Lorsqu'il est établi que l'application active 6 a pour objet la recherche d'informations via le réseau R (navigation Web), le module il initialise une communication vers le serveur de collecte S à l'aide de la primitive "connect" si cette communication n'est pas déjà établie. Les données de mesure 30 fournies par le décodeur de requêtes 9 sont ensuite envoyées au serveur S au vol. La connexion entre le module 11 et le serveur S est rompue lorsque toutes les "sockets" ouvertes par l'application 6 auront été fermées.

[0037] On a vu que la même entité peut gérer plusieurs serveurs de collecte S reliés au réseau R, différenciés par leur adresse IP et/ou leur port TCP. L'adresse IP et le port TCP du serveur S auquel doit se connecter le module 7 équipant un terminal particulier sont alors spécifiés de façon dynamique par le module 10 de gestion de la configuration. Ceci permet d'assurer un partage de charge entre les différents serveurs de collecte.

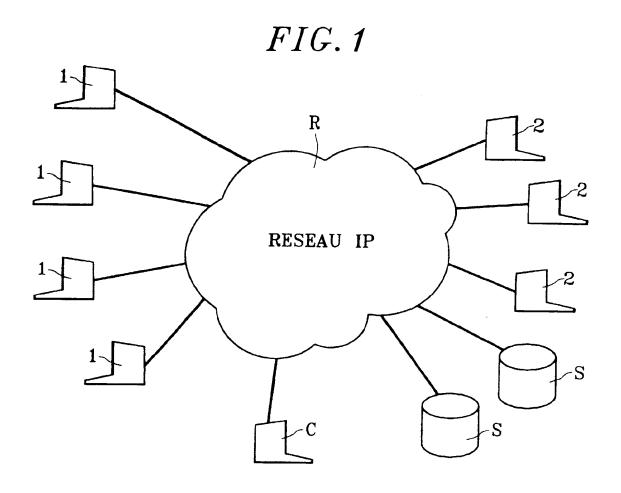
#### Revendications

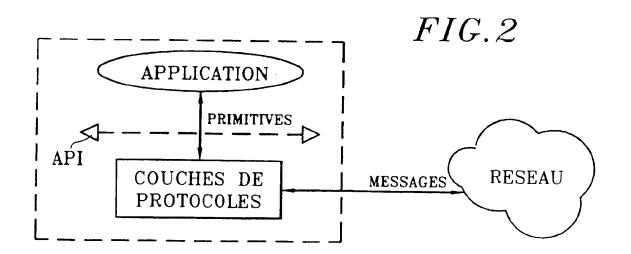
1. Procédé pour évaluer les consultations de sites (2)

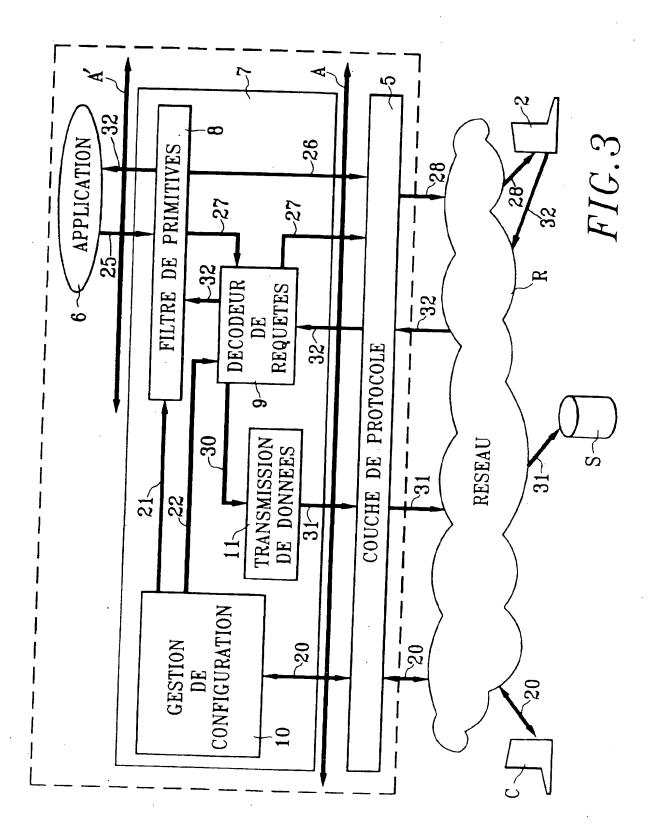
ou de canaux de diffusion accessibles par un réseau de télécommunications (R), dans lequel on relie au réseau au moins un serveur de collecte (S), et on équipe de moyens de mesure un ensemble de terminaux informatiques reliés au réseau, les moyens de mesure (7) dont sont équipés un terminal dudit ensemble étant agencés pour recevoir des messages d'accès émis par le terminal vers le réseau, pour retransmettre ces messages sur le réseau, pour détecter si chacun de ces messages concerne l'un des sites dont les consultations sont à évaluer et pour adresser des données de mesure au serveur de collecte par l'intermédiaire du réseau en réponse à la détection d'au moins un message concernant l'un desdits sites, le serveur de collecte assurant un traitement des données de mesures reçues pour estimer les nombres de consultations des sites.

- Procédé selon la revendication 1, dans lequel les moyens de mesure (7) dont sont équipés un terminal (1) évaluent la qualité de communication lors d'un échange de messages entre le terminal et l'un desdits sites par l'intermédiaire du réseau, et incluent dans les données de mesure adressées au serveur de collecte (S) des informations sur la qualité évaluée.
  - 3. Module de détection de messages et de transmission d'informations, à installer dans un terminal informatique (1) raccordé à un réseau de télécommunications (R), et comprenant:
    - un filtre de primitives (8) recevant des messages émis vers le réseau par des applications (6) exécutées dans le terminal, pour identifier ceux de ces messages qui se rapportent à des requêtes d'accès à certains sites (2) accessibles par le réseau, et pour transmettre les autres messages vers le réseau;
    - un décodeur de requêtes (9) recevant les messages de requête d'accès identifiés par le filtre de primitives (8), pour analyser ces messages et obtenir ainsi des données de mesure, et pour transmettre ces messages vers le réseau; et
    - des moyens de transmission de données (11) pour transmettre les données de mesure obtenues par le décodeur de requêtes à un serveur de collecte (S) accessible par le réseau.
  - 4. Module selon la revendication 3, comprenant en outre des moyens de gestion de configuration (10) pour obtenir d'un serveur de configuration (C) et/ou du serveur de collecte (S), via le réseau (R), des paramètres de configuration du filtre de primitives (8) et du décodeur de requêtes (9).

55









### RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande EP 99 40 1074

atégorie		Citation du document avec indication, en cas de besoin, des parties pertinentes Revendication concernée		CLASSEMENT DE LA DEMANDE (Int.Cl.6)	
D,Y	US 5 732 218 A (BLA) 24 mars 1998 (1998-	ND ET AL)	1,2	H04L29/08	
D,Y	WO 98 10349 A (NUWE 12 mars 1998 (1998- * page 42, ligne 23 figure 5C *	 B) 03-12) - page 43, ligne 20;	1-4		
Y	l avril 1992 (1992- * colonne 2, ligne		3,4		
A	data on the WWW* COMPUTER NETWORKS Al vol. 29, no. 8-13, septembre 1997 (199 XP004095329 * page 1350, colonn	ch of reliable usage  ND ISDN SYSTEMS,  7-09), page 1343-1355  e de droite, ligne 37 -  de droite, ligne 17 *	1-4	DOMAINES TECHNIQUES RECHERCHES (Int CI.6) H04L G06F	
	ésent rapport a été établi pour tou				
,	LA HAYE	Date d'achèvement de la recherche	0   5+~	Examinateur	
X ; part Y ; part	ATEGORIE DES DOCUMENTS CITE iculièrement pertinent en consideration de la même carégorie	E : document de bre date de dépôt ou	pe à la base de l'i evet antérieur, ma la près cette date ande	ais publié à la	

#### ANNEXE AU RAPPORT DE RECHERCHE EUROPEENNE RELATIF A LA DEMANDE DE BREVET EUROPEEN NO.

EP 99 40 1074

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de rechorche européenne visé ci-dessus.

Les dits members sont contenus au fichier informatique de l'Officeeuropéen des brevets à la date du Les renseignements foumis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

14-09-1999

Date de publication	Membre(s) de la unille de brevet(s)		Date de publication		cument brevet di apport de recher	
···	N	AUCI	24-03-1998	A	5732218	US
26-03-199 12-03-199 03-03-199 14-10-199	4243797 A 2246746 A 1209891 A 0870234 A	AU CA CN EP	12-03-1998	A	9810349	WO
18-03-199 04-01-199 18-04-199 13-09-199	0474932 A 69114805 D 69114805 T 5347524 A	EP DE DE US	01-04-1992	A	0478175	EP

EPO FORM POSE

Four tout renselgnement concernant cette annexe :voir Journal Officiel de l'Office européen des brevets, No.12/82

THIS PAGE BLANK (USPTO)